REMARKS

This communication is in response to the Office Action mailed on May 20, 2004. Before addressing the rejections made in the Office Action, applicant respectfully requests a drawing correction to FIG. 5. In particular, a double arrow has been indicated between the internet 205 and server 204. This double arrow was inadvertently omitted. Support for this drawing correction is found in the Specification on page 15, lines 1-4 wherein it is stated that the server 204 is connected and separately addressable through network 205. Entry of the replacement sheet is respectfully requested.

The Office Action first reports that claims 1-5, 7 and 11 and 18 were rejected under 35 U.S.C. 102(e) as being anticipated by Ladd et al. (U.S. Pat. 6,269,336). In particular, the Office Action reports that Ladd et al. disclose a computer readable medium including instructions readable by a computer that performs the steps including receiving data over network indicative of input at a client device and an indication of a grammar to be used with the data indicative of the input in order to perform recognition (col. 11, ll. 37-49 or col. 14, ll. 18-42); and sending data indicative of recognition results for the input to a remote location on the network (col. 8, ll. 55-67).

It is believed that Ladd et al. disclose in the most relevant embodiment illustrated in FIG. 3 a system 200 that allows users of communication devices indicated at 201, 202, 203 and 204 to access information stored on content providers 208 and 209 using a communication node 212 (col. 5, ll. 12-38). It is believed a summary of some aspects of the system 200 is provided at col. 11, ll. 25-63, wherein response to voice inputs from the user or DTMF tones, presumably using one of the connection devices 201-204, the voice browser 250 can navigate to a designation or content provider 208, 209. After the voice browser 250 is connected to an information source, the information source provides information that can include text content, mark-up

language documents or pages, non-text content, dialogs, audio sample data, recognition grammars, etc.

Based on the information collected, the voice browser 250 allows interactive voice applications. FIGS. 5A-5C illustrate a flow diagram for providing an interactive voice application. This procedure is discussed at col. 13, ll. 66-col. 15, ll. 59. Voice browser 250 accesses and uses a voice response unit server 234 having a text-to-speech converter 252 and a speech recognizer 254. One of the steps of voice browser 250 must undertake in order to perform a speech recognition is to determine whether a pre-determined grammar exists for the user input which is described at col. 14, ll. 18-42. Since the voice browser 250 is used to allow information to be received from the content provider 208/209, it is believed that the grammars appropriate for obtaining the information is closely associated with the content providers 208, 209 or contained in the markup pages from mark-up language servers 251, 257. Once the grammar has been established, the voice browser can then match the user input to the grammar in order to provide an interactive voice application (col. 14, ll. 41-42).

The inventions recited by the independent claims of the present application are patentably different from the system taught and suggested by Ladd et al. FIG. 5 of the present application illustrates a architecture 200 for web based recognition. As best summarized at p. 14, ll. 15-29, the architecture includes a client device 30, a web server 202 and a recognition speech server 204. When recognition is desired such as voice recognition on the client device for an application provided by the web server 202, the client device 30 may not be capable or powerful enough to perform voice recognition and as such can offload this task to the speech server 204. In particular, the client device 30 provides data indicative of the audio signals from the user as well as an indication of a grammar or language model to be used during speech recognition by the speech server 204. In other words, the speech server receives data indicative of

what the user has spoken as well as a grammar to perform recognition. The speech server 204 performs recognition and returns the results back to the client device 30 for local rendering if desired or appropriate. As discussed further on page 15, ll. 7-13, using the architecture described, authoring at the web server 202 can be focused on the application to which it is intended without the authors needing to know the intricacies of the speech server 204. Instead, the speech server 204 can be independently designed and connected to the network 205 and be updated and approved without further changes required at the web server 202.

Claim 1 has been amended to clarify the patentable differences of the present invention over the system described by Ladd et al. In particular, it is clarified that the network comprises a wide area network and the step of receiving data includes receiving input from a client device as well as an indication of a grammar from the client device to be used for recognition. Nowhere does Ladd et al. teach or suggest that the connection devices 201-204 that provide a user input also provide a grammar. Rather, it is believed that that grammar originates from the content provider 208-209 or the language markup servers 251, 257, which are apparently associated with the content providers 208-209.

In view of the foregoing, it is respectfully believed that independent claim 1 is allowable as discussed above. Dependent claims 2-3 and 7-10 recite further features of the invention recited by independent claim 1 and are believed to be separately patentable. It is respectfully noted that the Examiner has taken official notices that handwriting, gesture and visual recognition are well known in the art and that would have been obvious to one of ordinary skilled in the art at the time the invention to readily combine handwriting, gesture and visual recognition system of Ladd et al. thus, rendering these claims unpatentable. Applicants respectfully request the references be provided for handwriting, gesture and visual recognition that

would suggest or teach the inventions recited by these claims. As discussed above, its is believed that Ladd et al. does not teach or suggest the invention recited by claim 1, and therefore, the reference that would teach handwriting, gesture or visual recognition that would otherwise suggest the features of claims 8-10, in combination with the features recited in independent claim 1 would not render these claims unpatentable.

Independent claim 11 recites a method for speech recognition in a client/server network in a manner similar to that recited by claim 1. Applicant has amended claim 11 in a manner similar to that discussed above in claim 1 and it is believed that for the reasons discussed above, claim 11 is also now in condition for allowance. Dependent claims 12-15 recite further features of the invention and are believed separately patentable when combined with independent claim 11 and any intervening claims. Allowance of these claims is also respectfully requested.

Independent claim 16 recites "a computer readable medium having a markup language for execution on a client device in a client/server system, the markup language comprising instructions to unify at least one of recognition-related events, GUI events and telephony events of non-display, voice input based client device in a multimodal based client device or a web server interacting with each client device." Multimodal is a particular mode of entry that allows the user to use speech recognition in conjunction with at least a display in order to easily interact with the application running thereon. As stated in the Specification at p. 20, ll. 12-27, in this mode of data entry, the user is generally under control of when to select a field and provide corresponding information. In the example of FIG. 6 of the present application, a user can select a field and then provide speech input corresponding to that field. Again, the user is under control when to select a field and then to provide the corresponding speech input. Independent claim 16 in particular recites that a markup language executable on a client device includes instructions "to unify at least one of recognition-

related events, GUI events and telephony events on non-display, voice input based client device and a multimodal based client device for a web server interacting with each of the client devices."

The Office Action recites p. 5, ll. 1-11 for teaching the invention of claim 16. However, this portion of Ladd et al. appears to generally describe what information can be obtained and further states that system 200 "enables application developers to build applications for interactive speech applications using a markup language such as VoxML™ Voice Markup Language developed by Motorola, Inc." It is not understood how this passage can render claim 16 unpatentable for it does not teach or suggest the claimed invention. Accordingly, it is believed that claim 16 is allowable along with the dependent claims 17-19.

A petition for an extension of time is hereby requested. A charge authorization is included herewith for the extension fee.

The Director is authorized to charge any fee deficiency required by this paper or credit any overpayment to Deposit Account No. 23-1123.

Respectfully submitted,

WESTMAN, CHAMPLIN & KELLY, P.A.

By: _____

Steven M. Koehler, Reg. No. 36,188
Suite 1600 - International Centre
900 Second Avenue South
Minneapolis, Minnesota 55402-3319
Phone: (612) 334-3222 Fax: (612) 334-3312